REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	18 Nov. 98	1 Nov. 97 - 1 Nov. 98

4. TITLE AND SUBTITLE

Automate Information Extraction from Scan Data

5. FUNDING NUMBERS

6. AUTHOR(S)

Nurre, Joseph H.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Ohio University
Office of Research and Sponsored Programs
105 Research & Technology Center
Athens, OH 45701

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Logistics Agency
MMPRT Room 3135
8725 John J. Kingman Road, #2533
Fort Belvoir, VA 22060-6221

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The DDFG-T2-P3 Short Term Project is a continuation of the DDFG-T2-P3 Short Term Project. The goal of the two projects is the automated extraction of apparel measurements from scan data. The development of algorithms for three dimensional data analysis and processing proceeded to a new phase. The algorithms were field tested at the Marine Corps Recruit Depot in San Diego (MCRD-SD), in cooperation with the Cal Poly-Pomona Demo and the USMC. This required that robustness be added to the algorithms, where necessary. A new suite of software tools was also developed to manage the scan data. Furthermore, Ohio University provided technical support and man-power for the body scanning experiment. A CD is included with this report that presents the project's results in a PowerPoint presentation. Source code, for methods developed or refined, were released to Cyberware, Inc.

14. SUBJECT TERMS	15. NUMBER OF PAGES
3D Scanning, segmentation, apparel sizing	40
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

DLA-ARN Final Report
for

US Defense Logistics Agency

on

DDFG-T2/P5: ***AUTOMATE INFORMATION EXTRACTION FROM SCAN DATA***

Contract Number	SPO100-95-D-1014
Contractor	Ohio University
Delivery Order #	0002
Delivery Order Title	Automate Information Extraction from Scan Data
CDRL #	A005
CDRL Title	Technical Report, Final Report
Reporting Period	November 1, 1997 – November 1, 1998
Report Date	November 18, 1998
Name of PI's	Joseph H. Nurre
E-mail	nurre@bobcat.ent.ohiou.edu
Phone	614/593-0103
Fax	614/593-0007
Address	School of Electrical Engineering & Computer Science Ohio University Athens, Ohio 45701

Submitted to:

Apparel Research Network
Defense Logistic Agency
2800 South 20th Street
Philadelphia, Pennsylvania 19101-8419

Table of Contents

DDFG-T2/P5: *AUTOMATE INFORMATION EXTRACTION FROM SCAN DATA*

Executive Summary

The DDFG-T2-P5 Short Term Project is a continuation of the DDFG-T2-P3 Short Term Project. The goal of the two projects is the automated extraction of apparel measurements from scan data. The development of algorithms for three dimensional data analysis and processing proceeded to a new phase. The algorithms were field tested at the Marine Corps Recruit Depot in San Diego (MCRD-SD), in cooperation with the Cal Poly-Pomona Demo and the USMC. This required that robustness be added to the algorithms, where necessary. A new suite of software tools was also developed to manage the scan data. Furthermore, Ohio University provided technical support and man-power for the body scanning experiment. A CD is included with this report that presents the project's results in a PowerPoint presentation. Source code, for methods developed or refined, were released to Cyberware Inc.

1. INTRODUCTION

1. Objectives

The objective of T2P5 is to automate the analysis of 3D body scans for sizing. From data collection to uniform issuing, the entire clothing process for recruits can become electronic, resulting in both efficiency and accuracy. Data collection is the first link in the issuing process. Fewer mistakes up front results in lower overall costs. 3D body scanning also offers additional sizing information that is too expensive to collect using traditional measurement techniques. Software developed for the T2P5 project is a collaborative effort of Ohio Universtiy, Cyberware, Anthropology Research Project, Inc., and Beecher Research, Inc.

2. San Diego Marine Recruit Station Experiment

A scanning experiment was conducted at the San Diego Marine Recruit Station. Over 700 scans were taken, as a group of recruits were followed from their initial induction at the "night room", to their final fitting before graduation. In every aspect of a Marine's training, time is money. The training calendar for the recruits is rigidly enforced. The Marines bring into their San Diego Recruit Station between 300 (off peak) and 700 (on peak) recruits per week. The recruits are sized and fitted over two days in the morning. Using a scanner for sizing must take less than 4 hours to fit effectively into the training schedule.

The scanner hardware and sizing software preformed admirably through the experiment. Scanning required 17 seconds, and no hardware break downs occurred. Software processing took 30 seconds. The software was ~98% reliable with respect to software crashes. Recovery

from crashes was quick, and did not slow the recruit processing time. Total processing time for the recruits, including posing and comparison measurements by Anthropologists was one minute.

At a minute processing time, two scanners could quickly and efficiently process all recruits. Furthermore, software improvements have already resulted in a significant reduction in computer time. The experiment proved conclusive the practicality of the scanner process. Ohio University provided over 450 man hours of direct support for the experiment. Two technical support personnel were required for five scanning sessions. Only one technician from Ohio was required for the sixth and final session, because Cyberware had sent additional manpower to dismantle the scanner. All support personnel attended a planning meeting at Cyberware.

3. Data Analysis Software

Over 300,000 three-dimensional data points are picked up from the scanner, resulting in 900,000 numerical values. There are two primary challenges to developing sizing software. First, it must select among 300,000 data points to find only a specific few points that are important for apparel measurement. For example, only two points are needed to measure across the subject's shoulders.

Second, it must be able to consistently find these measuring points, despite the wide variety of shapes that humans come in. It is in fact easy to identify measurements on a select few subjects. But once you get into hundreds of subjects the issue becomes difficult.

The selectivity process is broken down into several steps. Each step discovers new information about the scan.

- First, noise in the scan data is removed. The data captured from the scanner contains noise points, known as curl. These are points that occur above the surface of the subject. A noise removal technique was developed that eliminates the curl. Removing curl is important for accurate measurements.
- Second, the subject is segmented. Segmentation is performed using computational geometry, imaging algorithms and a cusp locator algorithm. The simplest description of a cusp would be two arcs that meet at a point. A cusp then is like an arrow pointing the way between cylinders, such as the legs. When there is a large separation between the legs, the cusp points out the separator. As the legs come together, there is still a cusp or arrow pointing out the bisector between the legs. In this way, the cusp algorithm uses the geometry of slices to separate the different extremities of the subject.
- Third, with information on body segments and orientation, data selection can be limited to a specific area. Further processing of the subset can then discover needed points of measurement. To find shoulder and neck points for example, the head/upper torso area is processed. A computer can not "see" a subject's profile, however, it can calculate slope and distance. Slope data can be used to find the shoulders.
- Finally, measurements are taken in a fashion similar to that of a tailor. A path is created along the subject's surface, in order to simulate a tape measure.

The software crashed less than 2% of the time during San Diego Marine Recruit Station Experiment. Several of these crashes have been isolate, and correction have been made to the code. The Ohio University segmentor required nearly 30 seconds to operate during the

experiment. It now operates in only 10 seconds. Twenty-three measurements have been implemented requiring just over a minute to calculate.

4. Information Management

Scanning large numbers of recruits results in hundreds of large data sets. This can create challenges in organizing the scans, extracting information from the scans, moving the scan information and storing the scan files. These challenges were met successfully by creating several software tools.

The scan information editor is a tool created to organize the data. The organization of scans follows a progression from a large (over 700) scan set to a single piece of information related to a single subject. As the user processes the scans, the software developed at Ohio University keeps track of his progress and displays the total number of scans examined. The user may display information from a particular scan by selecting it from the list. The information pertaining to this scan will be displayed in the bottom of the tool.

Automatic information extraction tools were developed by Ohio University. The tools allow the user to select the scans from which to extract information and the measurements, used in the process. The output is flexible and can be used in a number of forums. It can be used internally with CyScan. It can be exported to a format readable by Excel. The output can also be directed to a cgi script readable file thereby allowing access via the Internet.

Compression management software was also developed by Ohio University. This was necessary because of the massive amounts of scan data exceeded hard drive space, and was expensive to transmit. A tool was developed to help the user organize and compress the proper

scan data. A list box displays all the scans in a given scope. The scope can be all the scans on a particular day, a particular month or the entire set. A display of the total scans, total scans compressed and total disk space used is provided. The user may select from the list box the scans to be compressed. Pressing the compress button compresses the selected scans. After compression new statistics relating to the compressed scans are given.

5. Conclusion

With the DDFG-T2P5 funding, significant progress was made towards successfully implementing a system that manages and extracts tailor measurements from 3D body data. Ohio University contributed to the planning and provided personnel support for the Marine experiment.

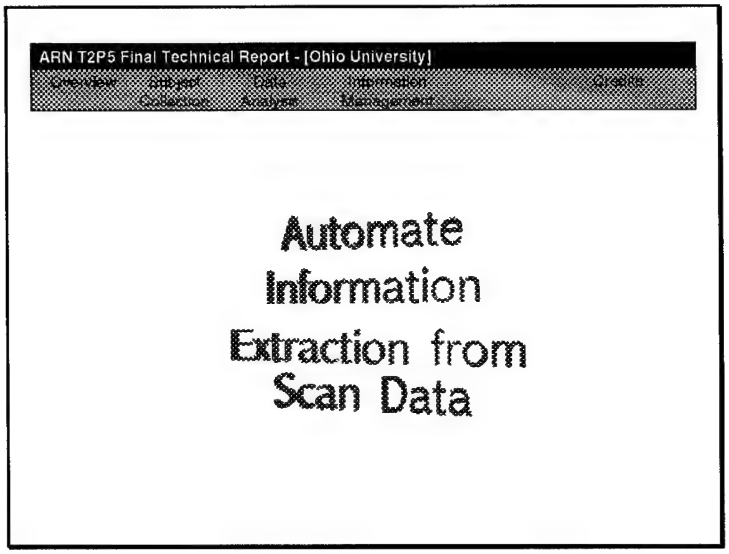
Based on the experiment, Ohio University improved the measurement code. Any software crashes due to Ohio algorithms were tracked and repaired. Segmentation techniques have been sped up three fold and now operate in only ten seconds. Data base software was developed to help manage the scan to issue process. Finally, a PowerPoint presentation was produced to document progress. The presentation required extensive macro programming and design, as well as, original movie animations that visualize and give meaning to complex ideas.

APPENDIX A

Hard copy of the T2P5 PowerPoint presentation.

(Four CDROM copies of the presentation were included with this report.)

Slide 1



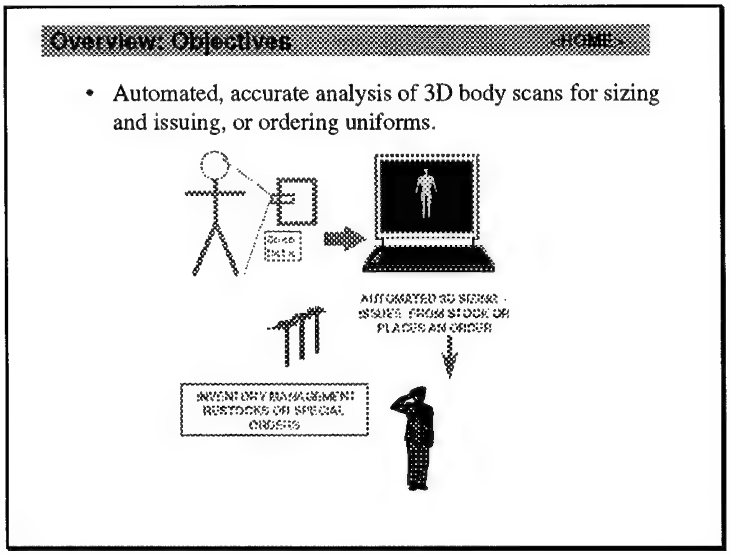
To emphasize the software aspects of the T2P5 project, this presentation was formatted to resemble a software package. The "Main Menu" on this title page leads you through the challenges and successes of the development effort. It is suggested that you start with the left most pull down menu and move to the right, clicking on all options. Clicking on the blue title bar will cause all menus to become visible. If at any time during the presentation, the screen goes white, try simply moving the mouse to return to the main menu.

Slide 2

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to 'fool' PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 3



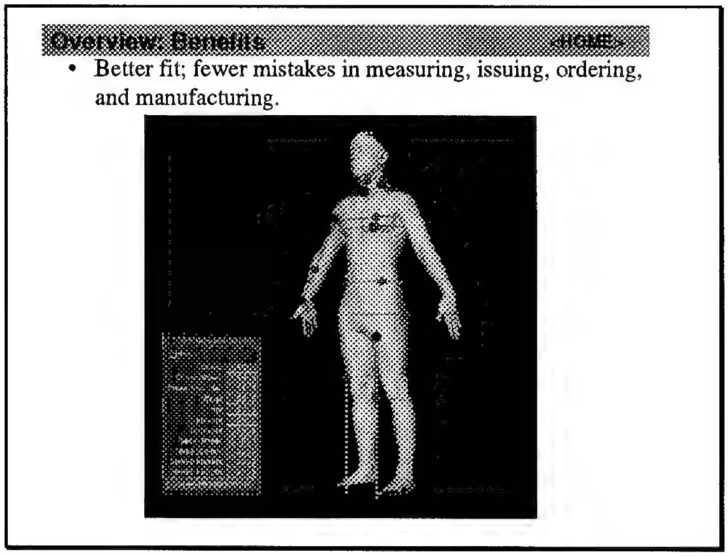
The objective of T2P5 is to automate the analysis of 3D body scans for sizing. From data collection to issuing, the entire process can become electronic, resulting in both efficiency and accuracy.

Slide 4

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to ‘fool’ PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 5



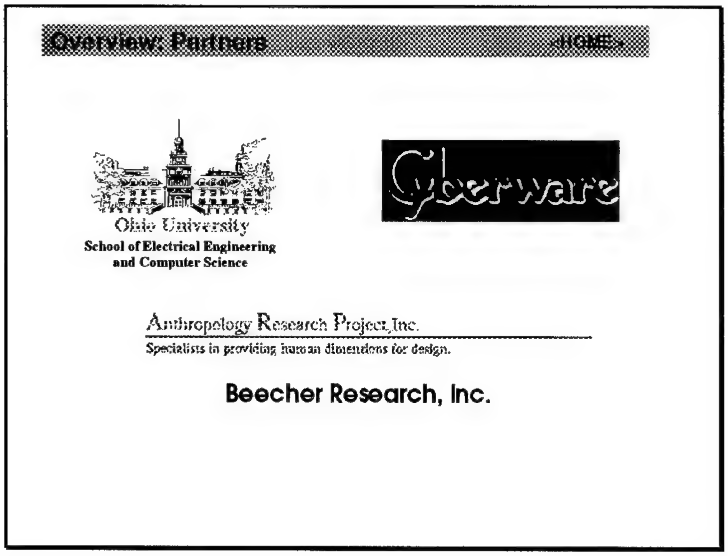
Data collection is the first link in the issuing process. Fewer mistakes up front results in lower overall costs. Furthermore, 3D body scanning offers additional sizing information that is too expensive to collect using traditional measurement techniques.

Slide 6

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to ‘fool’ PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 7



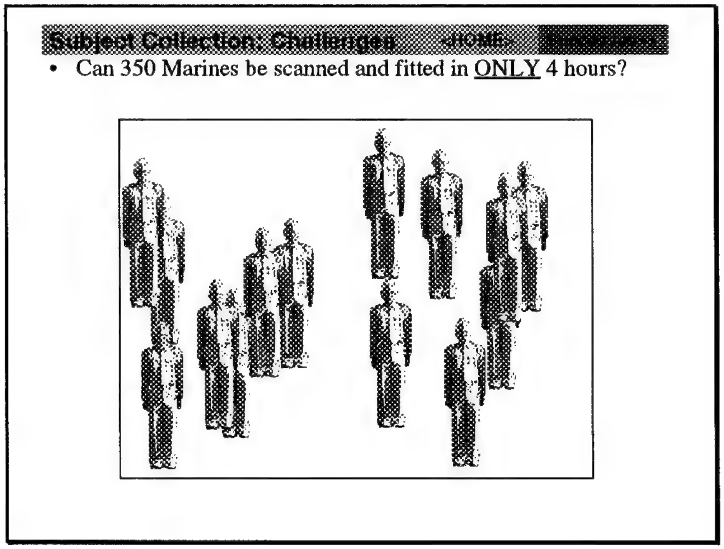
Software developed for the T2P5 project is a collaborative effort of four organizations. Ohio University provides the engineering tools in code, necessary to make fast and accurate measurements.

Slide 8

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to 'fool' PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 9



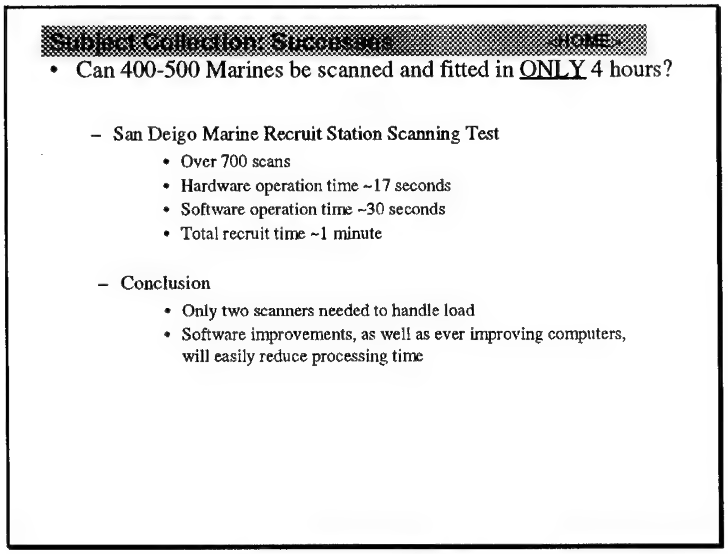
In every aspect of a Marine's training, time is MONEY. The training calendar for the recruits is rigidly enforced. The Marines will bring in between 300 (off peak) and 700 (on peak) recruits per week. The recruits are sized and fitted over two days in the morning. Scanning should probably take less than 4 hours to fit effectively into the schedule. To test the reliability of the hardware and software, a scanning experiment was conducted at the San Diego Recruit center. Ohio University provided two technical support personnel for nearly every scanning session.

Slide 10

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to ‘fool’ PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 11



As mentioned, a scanning experiment was conducted at the San Diego Recruit center. Over 700 scans were taken, as a group of recruits were followed from their initial induction at the "night room", to their final fitting before graduation. The hardware/software system setup preformed admirably. Scanning required 17 seconds, and no hardware break downs occurred. Software processing took 30 seconds. The software was ~98% reliable with respect to software crashes. Recovery from crashes was quick, and did not slow the recruit processing time. Total processing time for the recruits, including posing and comparison measurements by an Anthropologist was one minute.

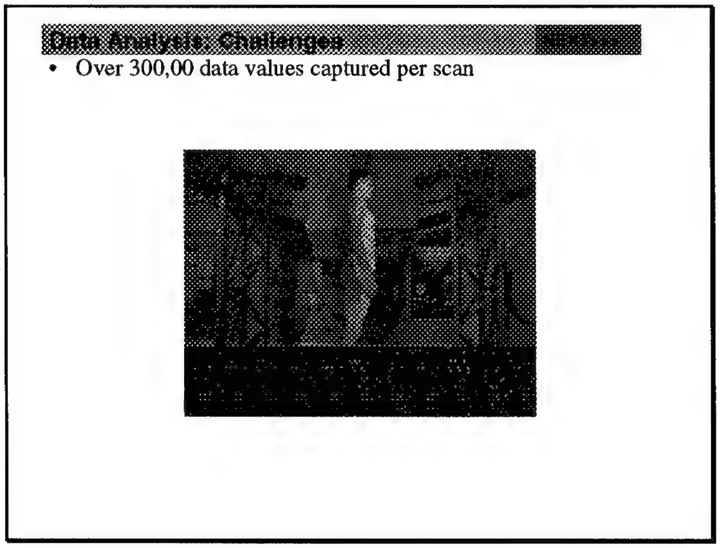
At a minute processing time, two scanners could quickly and efficiently process all recruits. Software improvements have already resulted in a significant reduction in computer time. The experiment proved conclusive the practicality of the scanner process.

Slide 12

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to ‘fool’ PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

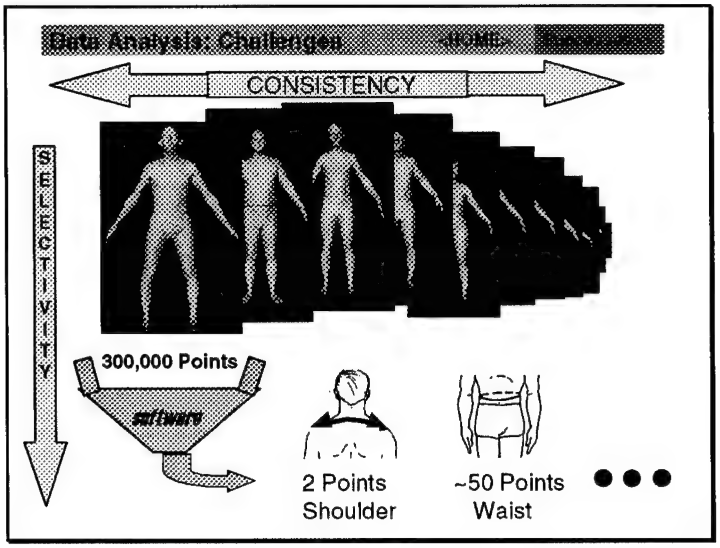
Slide 13



This is the Cyberware full body scanner. The subject stands in the middle of four scanning heads. The scanner operates by projecting a laser light that can be seen as a red line along the subject. The laser light projects from the middle of the scan head and falls on the subjects surface where it is picked up as two images. One image is taken near the top of the scan head, one is at the bottom. It takes about 17 seconds for the scanner to operate.

Over 300,000 three-dimensional data points are picked up from the scanner, resulting in 900,000 numerical values. Imagine putting 900,000 values into a spread sheet.

Slide 14



There are two primary challenges to developing software to work with this data. First, it must SELECT among the 300,000 data points to find only a specific few points that are important for apparel measurement. For example, only two points are needed to measure across the subject's shoulders.

Second, it must be able to CONSISTENTLY find these measuring points, despite the wide variety of shapes that humans come in. It is in fact easy to identify measurements on a select few subjects. But once you get into hundreds of subjects the issue becomes difficult.

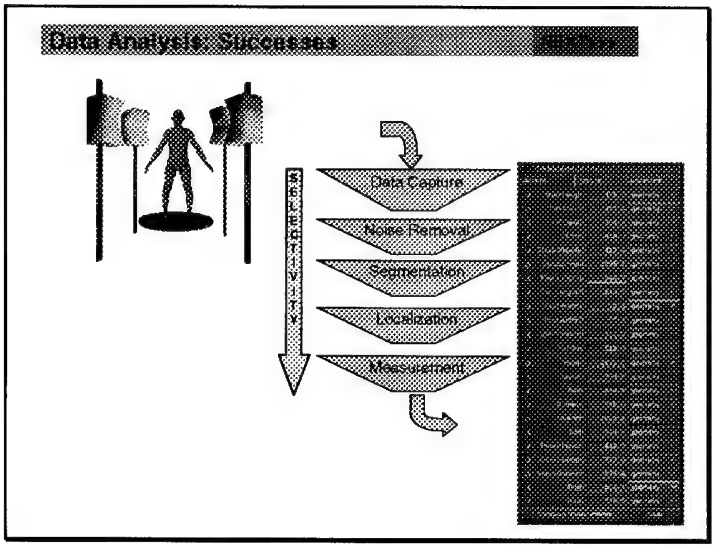
The software developed under ARN funding to present, is currently being tested on hundreds of scans. Preliminary results are very positive.

Slide 15

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to 'fool' PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 16



The scanner delivers an excessive amount of data, but no information. To SELECT the correct data points for measuring, information must be derived.

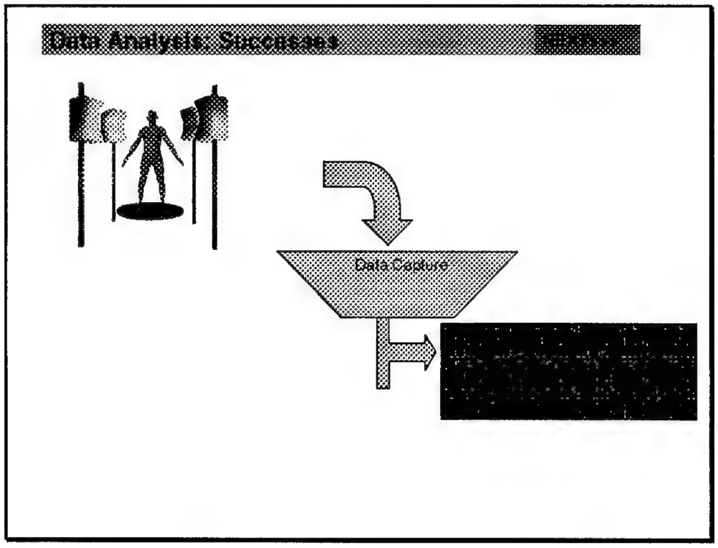
The SELECTIVITY process is broken down into several steps. Each step discovers new information about the scan.

- First, noise in the scan data is removed. All laser scanners generate a noise known as curl. Removing curl is important for accurate measurements.
- Second, the subject is segmented. Segmentation is performed using computational geometry and imaging algorithms. Some of the algorithms are new and were discovered under ARN funding.

- Third, with information on body segments and orientation, data selection can be limited to a specific area. Further processing of the subset can then discover needed points of measurement.

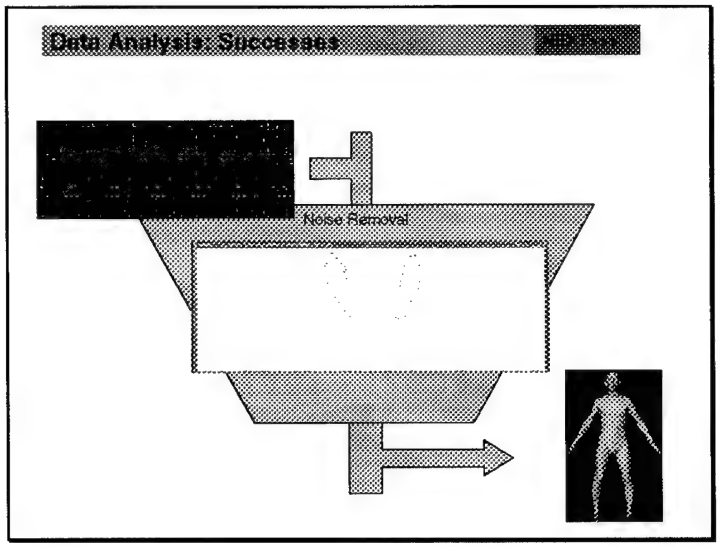
- Finally, measurements are taken in a fashion similar to that of a tailor.

Slide 17



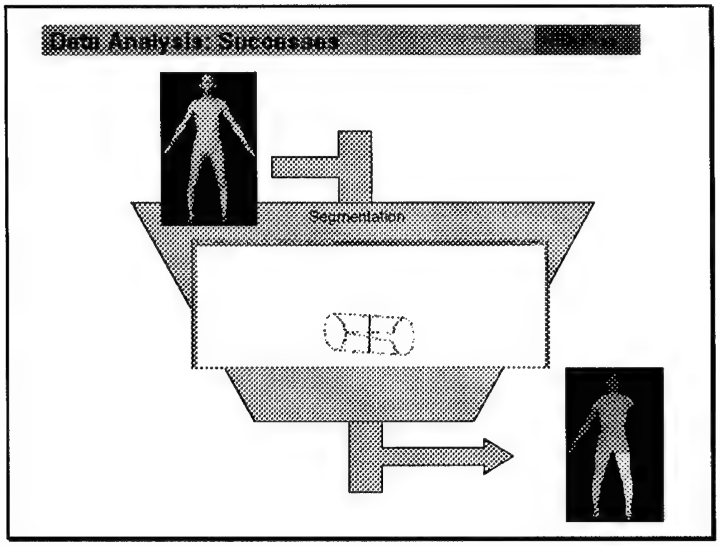
Over 300,000 three-dimensional data points are picked up from the scanner, resulting in 900,000 numerical values. It takes about 17 seconds for the scanner to operate.

Slide 18



The data captured from the scanner contains noise points, known as curl. These are points that occur above the surface of the subject. A noise removal technique was developed that eliminates the curl. In this slide, data is seen along horizontal slices. The slices start at the feet and end with the head. In each slice, data found to be noise is highlighted and removed.

Slide 19



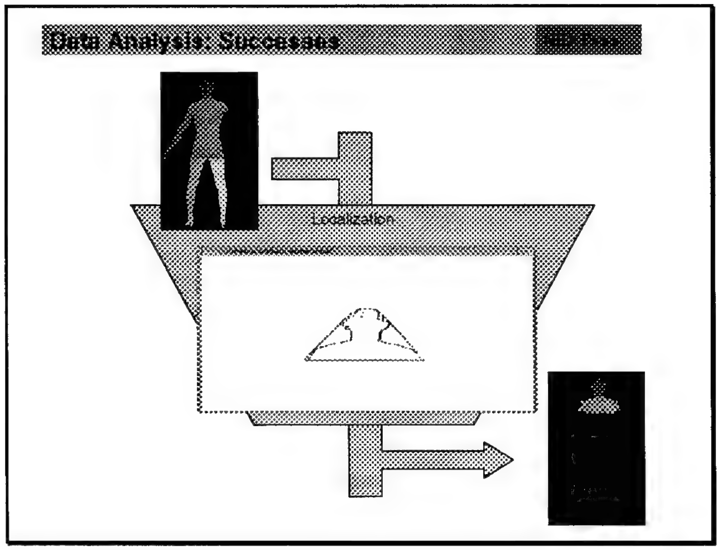
A cusp locator algorithm segments the body. The simplest description of a cusp would be two arcs that meet at a point. A cusp then is like an arrow pointing the way between the legs. When there is a large separation between the legs, the cusp points to the separator. As the legs come together, there is still a cusp or arrow pointing out the bisector between the legs. In this way, the cusp algorithm uses the geometry of slices to separate the different extremities of the subject.

This video clip demonstrates the cusp finder as it works its way up the legs. Cusps are located from the front and back of the legs. Even as the legs merge, a bisector can be found which separates the leg data.

A similar merging problem exists in data points between arms and chest. Notice that the segmentor continues to generate a bisector between the arm and the chest even after the arm and

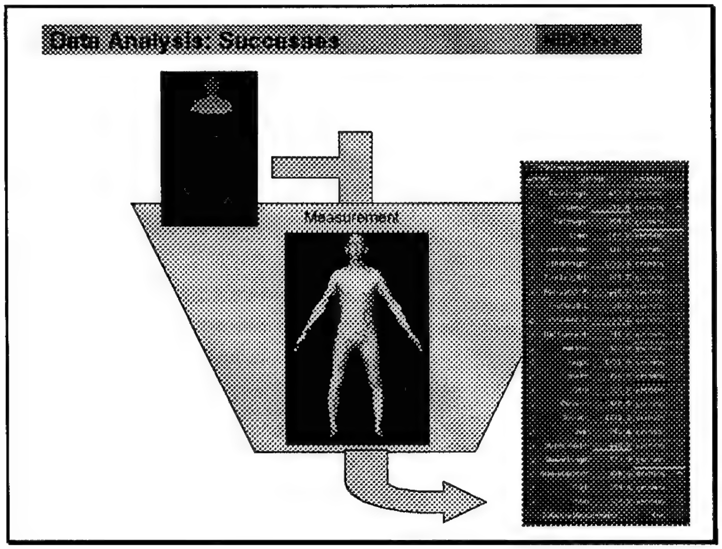
the chest data have merged together. The lines present towards the end of the process is a relaxation algorithm at work. A bisector is taken from the front and back of each slice. If these two cusps do not match up, then a relaxation algorithm is applied to force a match. At the end of each clip, when the extremity reaches the torso, the relaxation algorithm is matching up the front bisector with the back bisector to finish the separation.

Slide 20



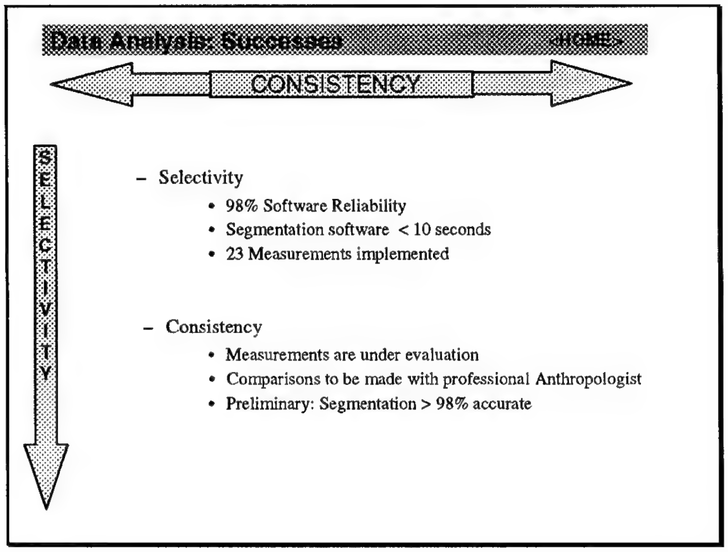
With information on body segments and orientation, data selection can be limited to a specific area. To find shoulder and neck points for example, the head/upper torso area is processed. A computer can not "see" the profile, however, it can calculate slope and distance. Notice the points in the video clip being classified as blue or white points, based on distance. Slope data can be used to find the shoulders.

Slide 21



Finally, measurements are taken from the scan data. A paths is created along the subject's surface, in order to simulate a tape measure.

Slide 22



This slide presents a summary of the software development.

Selectivity:

The software crashed less than 2% of the time at San Diego. This figure has since been improved. The Ohio University segmentor required nearly 30 seconds in San Diego. It now operates in only 10 seconds. Twenty three measurements have been implemented requiring just over a minute to calculate.

Consistency:

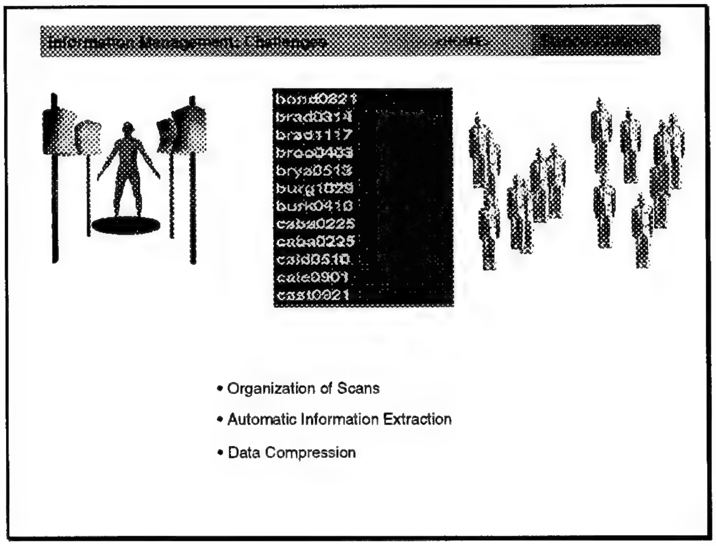
Measurement consistency is under evaluation.

Slide 23

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to ‘fool’ PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 24



Scanning large numbers of recruits results in hundreds of large data sets. This can create challenges in organizing the scans, extracting information from the scans, moving the scan information and storing the scan files.

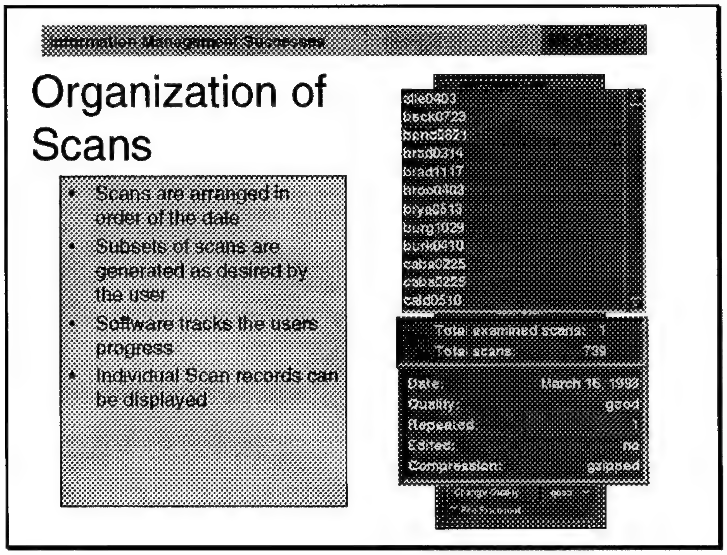
These challenges were met successfully by creating several software tools. One tool organizes the scans so that it may be processed efficiently. Another tool automatically extracts information from each scan so that the information may be used more efficiently. Yet another tool organizes a compression scheme, giving the operator the ability to greatly reduce the size of scanned files.

Slide 25

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to 'fool' PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 26

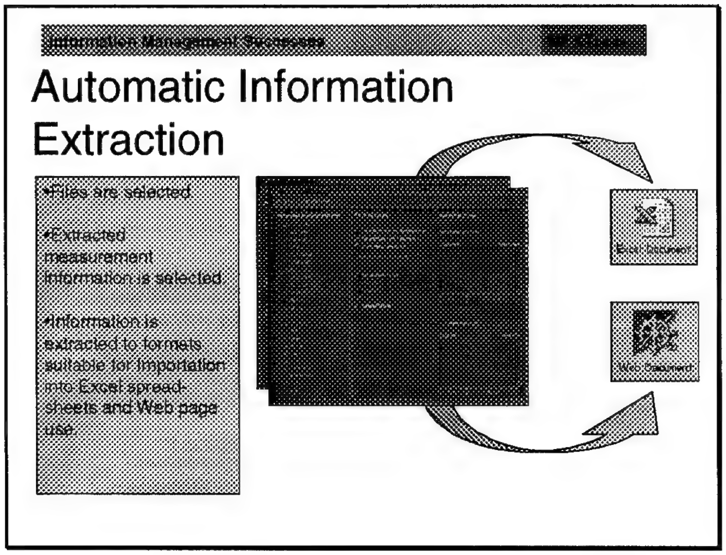


The scan information editor is a tool created to organize the data. The organization of scans follows a progression from a large set (over 700) scan to a single piece of information related to a single subject. (Clicking the mouse once will magnify the large data set.)

As the user processes the scans, the computer keeps track of his progress and displays the total number of scans examined. (Clicking the mouse a second time will magnify the statistics related to the users progress.)

The user may display information from a particular scan by selecting it from the list. The information pertaining to this scan will be displayed in the bottom of the tool. (The final mouse click will magnify the section of the GUI that displays information relative to an individual scan.)

Slide 27



Automatic information extraction tools were developed.

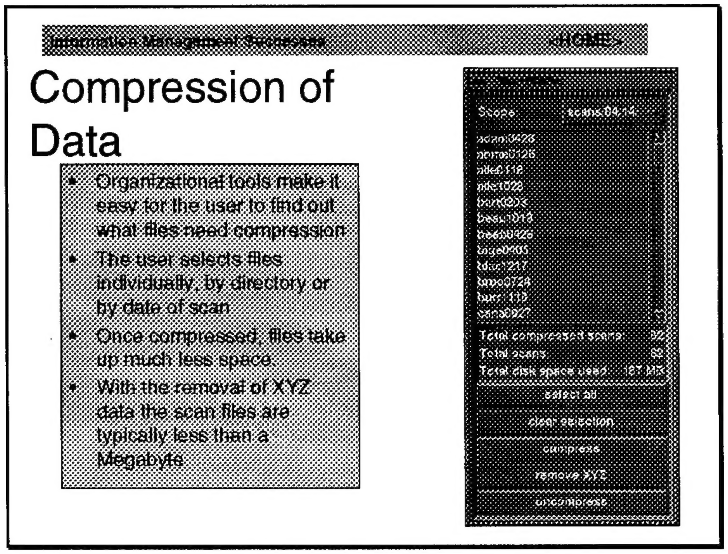
The tools allow the user to select the scans from which to extract information from and the measurements to use in the process.

The output is flexible and can be used in a number of forums. It can be used internally with CyScan. It can be exported to a format readable by Excel. The output can also be directed to a file which can be used by a cgi script thereby allowing access via the Internet.

(Clicking once causes a display of information going to an Excel file.)

(Clicking a second time causes a display of information going to an Web page document.)

Slide 28



Compression utilities were developed due to the massive amounts of data versus limitations in hard drive space, and costs of transmitting large amounts of data.

A tool was developed to help the user organize and compress the proper scan data.

A list box displays all the scans in a given scope. The scope can be all the scans on a particular day, a particular month or the entire set. A display of the total scans, total scans compressed and total disk space used is provided. (Clicking the mouse once will narrow the scope from the entire set to a single day.)

The user may select from the list box the scans to be compressed. Pressing the compress button compresses the selected scans. (Clicking the mouse a second time will simulate the pressing of the compress button.)

After compression new statistics relating to the compressed scans are given. (Clicking the mouse again will show the new statistics results.)

Slide 29

THIS PAGE IS INTENTIONALLY LEFT BLANK.

It is required to 'fool' PowerPoint into acting more like a program with menus. Moving the mouse on this page, should bring you back to the main menu.

Slide 30

Credits HOME

Joseph H. Nurre PhD

Jeff S. Collier BSEE, MS

Erick A. Lewark BSEE, MS